Nested Regression Based Optimal Selection (NRBOS) of Rational Polynomial Coefficients

Long Tengfei, Jiao Weili, and He Guojin

Abstract

Although the rational function model (RFM) is widely applied in photogrammetry, the application of terrain-dependent RFM is limited because of the requirement for numerous ground control points (GCPs) and the strong correlation between the coefficients. A new method, NRBOS, based on nested regression was proposed to select the optimal RPCs automatically and to gain stable solutions of terrain-dependent RFM using a small amount of GCPs. Different types of images, including QuickBird, SPOT5, Landsat-5, and ALOS, were involved in the tests. NRBOS method performed better than conventional methods in estimating RPCs, and even provided a reliable solution when less than 39 GCPs were used. Additionally, the test results showed that the simplified RPCs are almost as accurate as the vendor-provided RPCs. Consequently, in favorable situations such as when the orientation parameters of the satellite are not available or are not sufficiently accurate, the proposed method has the potential to take the place of the regular terrain-independent RFM.

Introduction

The rational function model (RFM) with 78 rational polynomial coefficients (RPCs) is completely a mathematical model, which approximately describes the imaging process in photogrammetry and remote sensing. Terrain-independent RFM constitutes a comprehensive reparameterization of the rigorous sensor model, and is widely applied in high resolution image products. Terrain-dependent RFM, on the other hand, is hardly used because of the requirement for numerous observation data (ground control points (GCPs)), and the strong correlation between the coefficients. Actually, terrain-dependent RFM provides a useful approach to rectify remotely sensed images without knowing the position and orientation information of specific sensor. However, as 78 RPCs of the RFM are strongly correlated, stable, and precise solutions of the RPCs are difficult or even impossible to achieve (Lin and Yuan, 2008). The objectives of this research are to find a robust approach to estimate RPCs and to make use of the terrain-dependent RFM.

Many studies have been carried out on the topic of RFM during the recent decades. OGC (1999) has normalized the range of the image and object space coordinates of RFM to −1 to +1, and effectively enhanced the condition number of the normal equation matrix. Tao and Hu (2000) have studied the RFM comprehensively and have proposed to strengthen the solution of RFM using Tikhonov regularization and the L-curve method. Yuan and Lin (2008) have compared the results of several methods for solving RPCs including ridge trace method, L-curve method, empirical formula method, and generalized ridge estimate method; they have verified the validity of L-curve method. In addition, Levenberg-Marquardt method (Tao and Hu, 2001) and singular value decomposition method have been applied to solve RPCs (Fraser *et al.*, 2006). Among all these methods, the ridge estimation method (especially the L-curve method) is the most widely used approach. However, there are still some problems in solving RPCs using the existing methods. For example, ridge estimate is a biased estimate, which requires numerous GCPs to solve RPCs.

Solving RPCs is a problem of multiple regression analysis. The problem of multicollinearity results in an ill-posed normal equation, and the ordinary least squares (OLS) estimation does badly in achieving a stable and reliable solution. In order to solve the problem of multicollinearity, "variable selection (Draper *et al.*, 1966)" and "ridge estimation (Hoerl and Kennard, 1970)" are usually used to improve the OLS. Variable selection can simplify the original model by selecting a subset of variables from the original set of variables which give the most significant response to the regression, and therefore the multicollinearity of the model can be reduced after the variable selection (Guyon and Elisseeff, 2003). Both of the two methods have drawbacks. Variable selection provides interpretable model but can be extremely variable because it is a discrete process. Ridge regression is a continuous process that shrinks coefficients and hence is more stable. However, it does not set any coefficients to 0 and hence does not give an easily interpretable model (Tibshirani, 1996). In addition to the methods mentioned above, in statistics, some improvements of ridge estimation method (Bashtian *et al.*, 2011; Jurczyk, 2012; Kibria and Saleh, 2012; Park and Yoon, 2011) have been proposed in recent years. However, they still need a large number of observation data.

Variable selection is a non-deterministic polynomial (NP) problem whose search space is very large (Amaldi and Kann, 1998). During the past decades, hundreds of methods have been proposed to solve this problem, including genetic prediction, decision tree prediction, Bayes prediction, least square prediction, and support vector machine prediction (Guyon and Elisseeff, 2003), and all these methods need a huge computation. Greedy search strategy can greatly reduce the search space and thus improve the efficiency of the algorithm. The greedy search methods are mainly divided into three categories: forward selection, backward elimination, and stepwise regression analysis. Nonetheless, there are common disadvantages in these methods. Not all the possible combinations of variables are taken into account, and the results which greatly depend on the evaluation criterion are usually not the optimal solution. In addition, the variable transformation, such as principal component analysis, partial least

The Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China (wljiao@ceode.ac.cn).

Photogrammetric Engineering & Remote Sensing
Vol. 80, No. 3, March 2014, pp. 261–269.
0099-1112/14/8003–261

doi: 10.14358/PERS.80.3.261

squares (PLS) regression (Geladi and Kowalski, 1986; Lazraq *et al.*, 2003; Tobias, 1995), can be used to undermine the multiple-correlation, and a few independent variables that significantly respond to the regression are produced. However, the structure of the original regression model is also changed in the transformation, and the model may lack physical meaning. Besides, high fitting precision does not mean that the variable selection is appropriate although the PLS method performs better in fitting the observation data than stepwise regression (Chong and Jun, 2005).

Yuan and Cao (2011) have proposed an optimized method to select RPCs, which is actually a variable selecting method according to the evaluation criterion of multicollinearity and enhances the stability of RPCs' solution. As the multiple correlations between variables are complex, removing some relevant variables often leads to increased interpretation error of the model, as well as discards some useful information (Wang, 1999). Zhang *et al.* (2012) have proposed a new approach on optimization of the RFM, which has been utilized to the solution for terrain-independent RFM to alleviate the ill-posed problem and thus improve the stability of the estimated parameters. But the terrain-dependent RFM is not considered in their work.

In this paper, an automatic optimal selection of RPCs based on nested regression (NRBOS) is proposed, and the evaluation criterion of NRBOS is goodness of fit. The significant coefficients of RFM are selected step by step in the method, while the redundant coefficients are weeded out. The selected RPCs can be estimated using ordinary least square method.

The remainder of this paper is organized as follows: In the next Section, the linearized form of RFM is derived, followed by the proposed method of selecting optimal RPCs. Then, different types of remote sensing images are implemented with the three methods (OLS, L-curve, and NRBOS), and the experimental results are analyzed in detail. In the final Section, conclusions are drawn.

The Linearized Form of RFM

The rational function model (RFM) is a mathematical model, which describes the object-to-image space transformation. In order to improve the numerical stability of the equations, image coordinates and object coordinates are both normalized to the range of –1.0 to 1.0. The RFM is given in Equation 1:

$$\begin{cases} r = \dfrac{N_r(X,Y,Z)}{D_r(X,Y,Z)} \\ c = \dfrac{N_c(X,Y,Z)}{D_c(X,Y,Z)} \end{cases} \tag{1}$$

where, r and c are normalized coordinates of image points in image space, while X, Y, and Z are the normalized coordinates of ground points in object space,

$$\begin{aligned} N_r(X,Y,Z) &= a_0 + a_1X + a_2Y + a_3Z + a_4XY + a_5XZ + a_6YZ + a_7X^2 \\ &\quad + a_8Y^2 + a_9Z^2 + a_{10}XYZ + a_{11}X^3 + a_{12}XY^2 + a_{13}XZ^2 \\ &\quad + a_{14}X^2Y + a_{15}Y^3 + a_{16}YZ^2 + a_{17}X^2Z + a_{18}Y^2Z + a_{19}Z^3; \end{aligned}$$

$$D_r(X,Y,Z) = b_0 + b_1X + b_2Y + b_3Z + b_4XY + b_5XZ + \cdots + b_{19}Z^3;$$

$$N_c(X,Y,Z) = c_0 + c_1X + c_2Y + c_3Z + c_4XY + c_5XZ + \cdots + c_{19}Z^3;$$

$$D_c(X,Y,Z) = d_0 + d_1X + d_2Y + d_3Z + d_4XY + d_5XZ + \cdots + d_{19}Z^3;$$

a_i, b_i, c_i, d_i (i=0, 1, …, 19) are the rational polynomial coefficients (RPCs).

Generally speaking, b_0 and d_0 can be set as 1 after reduction of the fraction, and they are supposed to be 1 in this paper if there are no special explanations.

Although it is a nonlinear model, the RFM can be transformed into a linear model by a simple deformation, and the deformed linear model is shown as Equation 2:

$$\begin{cases} N_r(X,Y,Z) - rD_r(X,Y,Z) = 0 \\ N_c(X,Y,Z) - cD_c(X,Y,Z) = 0 \end{cases} \tag{2}$$

As the Equation 2 indicates, the row RPCs and the column RPCs are independent; it is possible to estimate the row RPCs and column RPCs separately. Taking the row RPCs for example, the first equation of Equation 2 is rewritten as Equation 3:

$$\begin{aligned} r = a_0 + a_1X + a_2Y + a_3Z + a_4XY + \cdots + a_{19}Z^3 \\ -b_1rX - b_2rY - b_3rZ - b_4rXY - \cdots - b_{19}rZ^3 \end{aligned} \tag{3}$$

When all the observation equations are ready, vector Equation 4 is obtained:

$$\mathbf{r} = \mathbf{X}\boldsymbol{\beta}_\mathbf{r} \tag{4}$$

where $\boldsymbol{\beta}_\mathbf{r} = [a_0, a_1, \cdots, a_{19}, b_1, b_2, \cdots, b_{19}]^T$, $\mathbf{r} = [r_1, r_2, \ldots, r_n]^T$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$, n is the number of the observation equations, $\mathbf{x} = [1, X_i, Y_i, \ldots, Z_i^3, -r_iX_i, -r_iY_i, \ldots, -r_iZ_i^3]$, and $i = 1,2,\ldots,n$.

In a similar way, the column RPCs equations can be derived.

According to Equation 4, solving RPCs is essentially a multiple linear regression problem, and the estimated value of $\boldsymbol{\beta}_\mathbf{r}$ may be obtained by ordinary least squares (OLS) method, as is shown as Equation 5:

$$\boldsymbol{\beta}_\mathbf{r} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{r} \tag{5}$$

where $\mathbf{X}^T\mathbf{X}$ is the coefficient matrix of the normal equation. Due to the strong correlation between the coefficients, $\mathbf{X}^T\mathbf{X}$ is usually ill-posed. In this case, direct inverse of the matrix may not yield accurate and stable results, and some ridge parameter is ordinarily brought in to strengthen the solution.

Nested Regression-based Optimal Selection (NRBOS) of RPCs

Goodness of Fit

In statistical analysis, regression lines vary from one regression model to another, and they more or less deviate from the observation data. Goodness of fit (Cameron and Windmeijer, 1997) of a statistical model describes how well the model fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model, including coefficient of determination and lack-of-fit sum of squares (Neter *et al.*, 1996). In this paper, the coefficient of determination is used as the statistical measure of goodness of fit.

The coefficient of determination is the proportion of the regression sum of squares in total variation, and it is given as Equation 6:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

where n is the number of observation, y_i is the i_{th} observation value, $\bar{y}$ is the mean of all the observation values, and $\hat{y}_i$ is the regression result of the i_{th} observation.

The value of R^2 is in the range of 0 to 1. The greater the value of R^2 is, the better the regression line approximates the observation data, and the more suitable the model (Pang, 2011).

Nested Regression

The concept of nested regression (Lin, 2008) is to divide the procedure of regression into several steps. One variable that fits the objective vector best is selected from the variable set in each step. When all the remained variables in the variable set are not significant for the objective vector, the procedure of nested regression is finished. Taking the statistical model (Equation 7) as an example, the procedure is shown as Figure 1:

$$\begin{aligned}\mathbf{y} &= b_0\mathbf{1} + b_1\mathbf{X}_1 + b_2\mathbf{X}_2 + \cdots + b_N\mathbf{X}_N + \mathbf{e} \\ &= b_0\mathbf{1} + \sum_{m=1}^{N} b_m\mathbf{X}_m + \mathbf{e}\end{aligned} \tag{7}$$

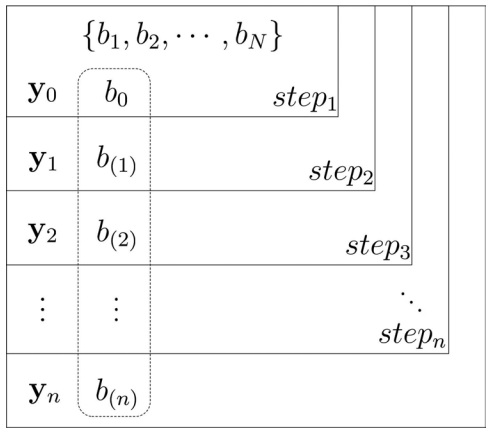

Figure 1. The concept of nested regression.

According to Figure 1, a variable set $\{b_1, b_2,\ldots, b_N\}$ is firstly collected and the first variable is b_0, while the initial objective vector y_0 comes from the observation data. In each step, a variable is picked out from the variable set $\{b_1, b_2,\ldots, b_N\}$. In this way, a new objective is derived from the selected variables and observation data. When the procedure is finished, the selected variables in each step, which are marked by the dashed box in Figure 1, constitute a new optimized variable set.

The full multiple linear regression model for Equation 4 is:

$$\begin{aligned}\mathbf{r} &= \mathbf{X}\boldsymbol{\beta}_\mathbf{r} + \boldsymbol{\varepsilon} \\ &= \beta_0\mathbf{1} + \sum_{i=1}^{N} \beta_i\mathbf{X}_i + \boldsymbol{\varepsilon}\end{aligned} \tag{8}$$

where $\mathbf{r}$, $\mathbf{X}$ and $\boldsymbol{\beta}_\mathbf{r}$ are the same as those in Equation 4, $\boldsymbol{\varepsilon}$ is the random errors, $\mathbf{X}$ can be rewritten as $\mathbf{X} = [\mathbf{1},\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_{38}]$, $\mathbf{1}$ is a column vector whose entries are 1, $\mathbf{X}_i$ is the i^{th} column vector of $\mathbf{X}$, such as:

$\mathbf{X}_1 = [X_1,X_2,\ldots,X_n]^{\mathrm{T}}$,
$\mathbf{X}_{38} = [-r_1Z_1^3, -r_1Z_2^3,\ldots,-r_nZ_n^3]^{\mathrm{T}}$, and so on

The full form of the result using ordinary regression method is:

$$\hat{\mathbf{r}} = \mathbf{X}\hat{\boldsymbol{\beta}}_\mathbf{r} \tag{9}$$

where $\hat{\boldsymbol{\beta}}_\mathbf{r}$ is the estimated value of $\boldsymbol{\beta}_\mathbf{r}$, and $\hat{\mathbf{r}}$ is the regression result.

RPCs Selection based on Nested Regression

Taking goodness of fit (coefficient of determination) as the evaluation criterion, a modified nested regression method is used to select the optimal rational polynomial coefficients. The procedure of the method is as following:

Step 1

Let the convergence thresholds $t_1 = 0.5$ / *scale* and $t_2 = 0.05$ / *scale*, where *scale* is the magnification for coordinates normalization. Let $k = 1$, $\mathbf{r}_k = \mathbf{r}$, and N_o is the number of observations, and N_v is the total number of RPCs. Then turn to Step 2.

Step 2

Together with $\mathbf{r}_k$, $\mathbf{X}_i$ (where $\mathbf{X}_i \neq \mathbf{X}_{(1)} \sim \mathbf{X}_{(k-1)}$) is successively used to build the linear regression models:

$$\mathbf{r}_k = \hat{\beta}_{k0}\mathbf{1} + \beta_{k1}\mathbf{X}_{(i)} + \boldsymbol{\varepsilon}_k \tag{10}$$

The coefficient of determination of each model is calculated respectively. Supposing the coefficient of determination of the model corresponding to $\mathbf{X}_{(k)}$ is the maximal one among all the coefficients of determination, Equation 11 is derived from the model corresponding to $\mathbf{X}_{(k)}$ using ordinary least square regression:

$$\hat{\mathbf{r}}_k = \hat{\beta}_{k0}\mathbf{1} + \hat{\beta}_{k1}\mathbf{X}_{(k)} \tag{11}$$

Then turn to Step 3.

Step 3

If $k \geq N_o$ or $k \geq N_v$, turn to Step 4; otherwise the residual vector is $\mathbf{v}_k = \mathbf{r} - \left(\sum_{m=1}^{k}\hat{\beta}_{m0}\mathbf{1} + \sum_{m=1}^{k}\hat{\beta}_{m1}\mathbf{X}_{(m)}\right)$, and the root mean square error is $\sigma_k = \sqrt{\mathbf{v}_k^T\mathbf{v}_k / N_o}$. If $\sigma_k < t_1$ and $|\sigma_k - \sigma_{k-1}| < t_2$, turn to Step 4; otherwise let $k = k + 1$, $\mathbf{r}_k = \mathbf{r}_{k-1} - \hat{\mathbf{r}}_k$, and turn to Step 2;

Step 4

Optimized variable set $\{\mathbf{X}_{(m)} | m=1,2,\ldots,k\}$ is obtained, and ordinary least squares method is used to regress the model (Equation 12), and the regression result is given as Equation 13:

$$\mathbf{r} = \beta_0\mathbf{1} + \beta_1\mathbf{X}_{(1)} + \beta_2\mathbf{X}_{(2)} + \ldots + \beta_k\mathbf{X}_{(k)} + \varepsilon \tag{12}$$

$$\hat{\mathbf{r}} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{X}_{(1)} + \hat{\beta}_2\mathbf{X}_{(2)} + \ldots + \hat{\beta}_k\mathbf{X}_{(k)} \tag{13}$$

Coefficients corresponding to $\mathbf{X}_{(m)}$ in $\boldsymbol{\beta}_\mathbf{r}$ are set as $\hat{\beta}_m$ (where $m = 1,2,\ldots,k$), and the other coefficients in $\boldsymbol{\beta}_\mathbf{r}$ are set as 0. So far, the procedure of calculating the RPCs is finished.

The flow chart of automatic optimal selection and solution of RPCs based on nested regression is shown as Figure 2.

In this procedure, the stop condition of variable selection is that the root mean square error is less than the threshold t_1 and the deviation of root mean square errors of the last two iterations is less than the threshold t_2. Typically, t_1 is set to 0.5 pixel, and t_2 is set to 0.05 pixel. In this case, if the estimation error is less than 0.5 pixel or the deviation is very small when a new variable is added, the reliable RPCs solution is obtained.

Tests and Analysis

Optimal Selection of RPCs Step-by-Dtep

A scene of a QuickBird P2AS image in Xi'an district of China, whose spatial resolution is 0.61 meters, was used in the test. The maximal difference in elevation of the image is about 650 meters, and 100 ground control points (GCPs) are evenly established using the originally provided RPCs on the actual

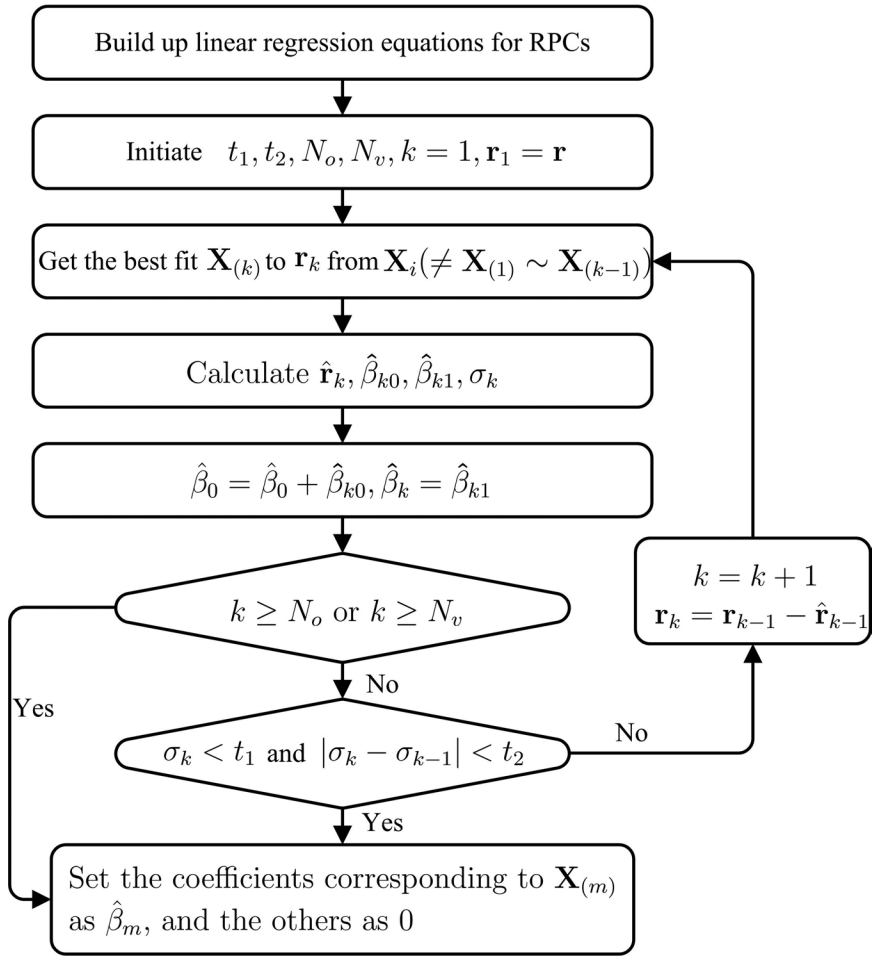

Figure 2. The procedure of RPCs selection and solution based on nested regression.

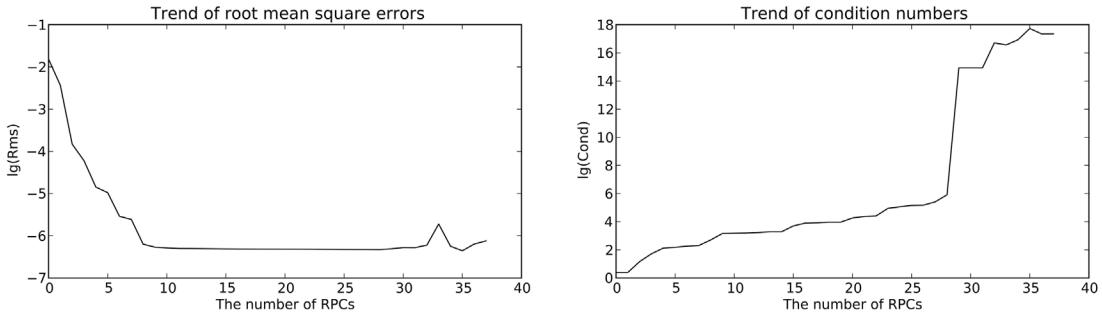

Figure 3. Root mean square error curve and condition number curve with the numbers of selected RPCs: (a) Root mean square error curve, and (b) Condition number curve.

surface of DEM (ASTER GDEM Version 2, with a spatial resolution is 30 meters). Taking the 100 GCPs as observation data (it is obviously terrain-dependent), the significant coefficients were selected from the 39 row RPCs according to the proposed method, and the root mean square error and condition number of normal matrix in each step of variable selection were noted. Then, the change curves of root mean square errors and condition numbers of normal matrix with the numbers of selected RPCs were plotted as Figure 3. In Figure 3a, the horizontal axis is the number of selected RPCs, and the vertical axis is the logarithms of root mean square errors. In Figure 3b, the horizontal axis is the number of selected RPCs, while the vertical axis is the logarithms of condition numbers. The purpose of using logarithms is to transform the data to similar orders of magnitude thereby enhance the discrimination of the data.

As can be seen from Figure 3, when there are few selected RPCs, the root mean square error is large, and the condition number of the normal matrix is small. With the increase of the number of selected RPCs, the root mean square error gradually reduces, while the condition number of the normal matrix gradually increases. However, when more than ten RPCs are selected, additional RPCs do not remarkably decrease the root mean square error any further; conversely, they may even increase the root mean square error. The main reason for the decrease of the accuracy is that the large condition number (more than 10^{16}) of the normal matrix makes the model

Table 1. The Condition Numbers of Normal Equation When Solving the RPCs With Three Different Methods

	OLS		L-Curve		NRBOS			
	Condition number		Condition number		Selected RPCs		Condition number	
	Row	Column	Row	Column	Row	Column	Row	Column
QuickBird	$4.92*10^{17}$	$2.25*10^{17}$	$9.88*10^{10}$	$1.02*10^{11}$	8	6	$1.68*10^{2}$	$3.28*10^{1}$
SPOT5	$1.61*10^{13}$	$1.50*10^{15}$	$1.68*10^{15}$	$3.15*10^{5}$	14	10	$2.21*10^{3}$	$7.10*10^{2}$
Landsat5	$9.63*10^{18}$	$3.96*10^{17}$	$3.94*10^{9}$	$1.53*10^{11}$	5	3	$7.00*10^{1}$	$4.03*10^{1}$

instable. Actually, the root mean square error of less than 10^{-5} pixels is small enough for the regression, although it slightly vacillates along the growth in number of parameters.

When few RPCs are selected, the coefficients are weakly correlated, and the condition number of the normal matrix is also small. Nevertheless the simple model with few RPCs does not approximate the observation data well, so the root mean square error is large. With the increase in the number of RPCs, the fitting accuracy gradually improves. Meanwhile, coefficients also become increasingly correlative, so the condition number of normal matrix increases rapidly, and the model becomes ill-posed and unstable.

Test of Different Types of Remote Sensing Images

In this study, three scenes of different types of remote sensing images, including QuickBird P2AS image, SPOT5 HRG 1A image, and a Landsat-5 TM L2 image, were used to verify the validation of the proposed method. For each remote sensing image, there were 120 evenly distributed ground control points. Three methods, including OLS, L-curve and NRBOS were applied to estimate the RPCs.

Comparison of Condition Number

In order to show the stability of NRBOS method, the condition numbers of the normal matrix of the three methods were compared. Table 1 shows the condition numbers of the normal equation when solving RPCs with three different methods and the number of RPCs selected from the row (column) function using nested regression method.

The following conclusions can be drawn from Table 1:

- For ordinary least squares, the condition numbers of the normal equation may be as great as 10^{18}. By using L-curve method, the condition numbers can be reduced to $10^{5} \sim 10^{10}$, but still indicate strong correlations between the RPCs. By using NRBOS method, only a few coefficients are selected from the 39 row (column) RPCs, and the condition numbers are reduced to $10^{1} \sim 10^{3}$.
- The numbers of selected RPCs vary from one image to another and that means the deformation varies from one image to another. In this test, the model of Landsat-5 L2 is the simplest one, and therefore it needs the least coefficients. Yet the model of SPOT5 HRG 1A is the most complex one, so the most coefficients are selected for the SPOT5 HRG 1A image.
- Generally, the numbers of RPCs selected from row function are more than those selected from column function for each type of image, which result from the different deformations in row direction and column direction. For the sensors with constant look-angle in the along-track direction (Landsat-5 L2 and SPOT5 HRG 1A), this indicates that the deformation in along-track direction is usually more complicate than that in the across-track direction. However, for very high-resolution, steerable satellites such as QuickBird, rows and columns invariably are not aligned to the along- and cross-track directions, we cannot hastily draw similar conclusions.
- The greater the condition numbers of the normal equation of OLS are, the fewer the numbers of selected RPCs in NRBOS method are. In fact, a small number of selected RPCs indicate that the deformation of the image is weak, therefore the 78 RPCs in full RFM is strongly correlated, and the normal equation in OLS will be severely ill-posed. However, to confirm this conclusion, more tests should be carried out in the future work.
- Normally, the number of RPCs selected from row function or column function is fewer than 20 for a regular satellite image, which indicates that more than half of RPCs are redundant.

Comparison of Accuracy

Then, 60 GCPs, 20 GCPs, and 10 GCPs were evenly selected from the original 120 GCPs successively, and OLS method, L-curve method and NRBOS were applied to calculate the RPCs, respectively. The GCPs were used to evaluate the fitting precision of the model, and the rest of the original GCPs were used as checkpoints (CPs) to evaluate the actual precision of the model. The test results are shown in Table 2. The quantities of optimal RPCs with different number of GCPs are almost the same as those shown in and the varieties are no more than two coefficients.

Several points can be obtained from the experiment (see Table 2):

- By selecting the optimal RPCs, the redundancy of the model can be reduced while the stability can be increased. When the GCPs are adequate (60 GCPs), both the OLS method and the L-curve method produce high fitting accuracy. However, the full RFM with 78 RPCs sometimes (in SPOT5 case) fails to fit the GCPs stably, and the accuracy of the CPs may be poor (more than 2 pixels). But the NRBOS method yields stable solutions. As only a subset of the RPCs is involved in the model, the fitting accuracy of GCPs is not as high as that in OLS method and L-curve method. Of course, one can select more RPCs by modifying the thresholds, resulting in higher fitting accuracy, but the additional RPCs do not necessarily improve the accuracy of the CPs.
- The NRBOS method can be used to estimate the RPCs when less than 39 GCPs are available. When the number of control points is less than 39, fitting errors in OLS method are 0, while the errors of CPs are exceptionally large. By adding some ridge parameters, the errors of CPs can be decreased to some extent, but the errors are still more than 10^{3} pixels. This is the so-called over-fitting phenomenon in multiple regression analysis, and in this case, both OLS method and L-curve method cannot get the correct RPCs. However, NRBOS method makes the number of the involved coefficients less than the number of control points, and therefore the reliable RPCs are obtainable.
- When the number of GCPs is less than the minimum number of required RPCs, the proposed method cannot get a stable solution. For instance, in the test using NRBOS method to estimate the RPCs of SPOT5 with ten GCPs, the errors of CPs in the row direction are more than 20 pixels (see Table 2). In fact, as the RFM for the SPOT5 HRG 1A needs at least 14 row RPCs (see Table 1), ten

TABLE 2. THE ACCURACY REPORT OF RPCS SOLUTIONS

Type of Image	NUM of Points	Type	RMS/pixel					
			OLS		L-Curve		NRBOS	
			row	column	row	column	row	column
QuickBird	60	GCPs	0.00	0.08	0.00	0.00	0.04	0.04
	60	CPs	0.11	0.16	0.07	0.10	0.05	0.03
	20	GCPs	0.00	0.00	0.00	0.00	0.05	0.43
	100	CPs	$3.85*10^3$	$1.23*10^4$	172	499	0.83	0.13
	10	GCPs	0.00	0.00	0.00	0.00	0.19	0.05
	110	CPs	$8.85*10^4$	$3.70*10^3$	$1.5*10^3$	$2.22*10^3$	1.94	0.15
SPOT5	60	GCPs	0.12	0.15	0.35	0.16	0.54	0.42
	60	CPs	2.64	0.68	1.25	0.87	0.69	0.79
	20	GCPs	0.00	0.00	9.69	16.0	0.32	0.36
	100	CPs	$1.21*10^4$	$7.05*10^4$	57.6	220	1.09	0.68
	10	GCPs	0.00	0.00	352	241	2.16	0.20
	110	CPs	$7.31*10^4$	$2.00*10^4$	$1.09*10^3$	$1.84*10^3$	27.7	1.80
Landsat5	60	GCPs	0.00	0.00	0.00	0.00	0.04	0.00
	60	CPs	0.68	0.45	0.68	0.45	0.68	0.45
	20	GCPs	0.00	0.00	0.00	0.00	0.02	0.00
	100	CPs	$3.38*10^3$	$1.58*10^4$	43.0	94.8	0.68	0.45
	10	GCPs	0.00	0.00	0.00	0.00	0.02	0.04
	110	CPs	$1.38*10^3$	$3.82*10^3$	470	476	0.69	0.45

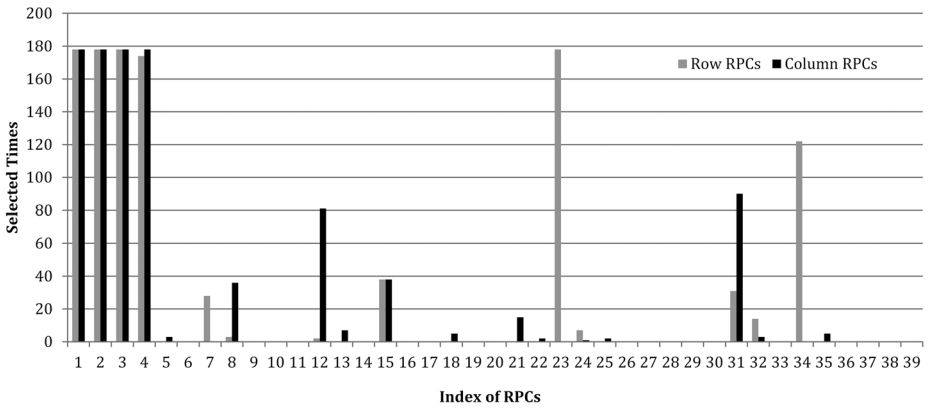

Figure 4. The frequency of each RPC to be chosen in the 178 tests.

GCPs fail to provide adequate observations to estimate the 14 row RPCs.

- When NRBOS method is used to estimate the RPCs, the accuracy in column direction is commonly better than that in row direction. For the sensors with constant look-angle in the along-track direction (Landsat-5 L2 and SPOT5 HRG 1A), this indicates that the deformation in along-track direction is usually more complicate than that in the across-track direction. This is related to the movement of the satellite.
- To estimate the RPCs, the number of necessary GCPs varies from one image to another. For the QuickBird and SPOT5 images, when the numbers of GCPs are reduced to ten from 60, the accuracy of CPs declines, although the selected RPCs are almost the same. However, 10 to 20 GCPs are normally sufficient for a regular satellite image by using the proposed method of NRBOS. In addition, for the Landsat-5 L2 image, the accuracy of CPs is almost unchanged when the number of GCPs reduces to ten. This result shows that the deformation of Landsat-5 L2 image can be fit by ten GCPs and get a stable imaging model. The additional GCPs do not notably improve the accuracy.

Potential for Substituting Regular Terrain-Independent RFM

In this subsection, 178 scenes of ALOS PRISM 1B2 (RPCs provided) images covering Jilin province of China were used, where the elevation range is from sea level to 2,650 m.

Get a simplified RFM

For each image, the RPCs provided by the imagery supplier (Takaku, 2011) were applied to establish 3D control grid consisting of five elevation layers each with 11 × 11 grid points, and then the method of NRBOS was used to estimate the reduced RPCs in terrain-independent way. In order to obtain high fitting precision, the convergence thresholds t_1 and t_2 were set as 0.005 / *scale* and 0.0005 / *scale*, respectively. The

selected RPCs were not exactly the same for every image, and the frequency of each RPC to be chosen was shown in Figure 4. All the row RPCs (from a_0 to b_{19} except for b_0) and the column RPCs (from c_0 to d_{19} except for d_0) were numbered in the horizontal axis consecutively.

According to Figure 4, we omitted those RPCs which were not frequently chosen (less than 40 times), and got a uniform model (Equation 14). As a matter of fact, the estimated values of those omitted RPCs are quite small (the absolute values are commonly less than 10^{-4}), and the fitting accuracy of the simplified RFM would not be affected much by excluding them from the model.

$$\begin{cases} r = \dfrac{a_0 + a_1X + a_2Y + a_3Z}{1 + b_3Z + b_{14}X^2Y} \\ c = \dfrac{c_0 + c_1X + c_2Y + c_3Z + c_{11}X^3}{1 + d_{11}X^3} \end{cases} \tag{14}$$

Precision of the Simplified RFM

For each image, the 5 × 11 × 11 grid points in previous subsection were used to estimate the simplified RFM. Then, the RPCs provided by the imagery supplier of each image were applied to establish 21 × 21 3D grid points on the actual surface of DEM (ASTER GDEM Version 2, with a spatial resolution is 30 meters). Finally the 21 × 21 grid points were used to check the precision of the simplified RFM (Equation 14). The test results obtained using ordinary least squares are provided in Table 3.

From Table 3, we can see:

- The accuracy in column direction is higher than that in row direction, which is similar to the previous test results.
- The accuracy in both directions can reach $10^{-2} \sim 10^{-3}$ pixels and the maximum errors are also less than 0.2 pixels. The result indicates that the simplified (Equation 14) can be used as a replacement of the regular terrain-independent RFM (Equation 1) for ALOS PRISM 1B2 data processing.

Application of the Simplified RFM

As the test result shows in the previous subsection, the simplified RFM can be used as the regular terrain-independent RFM, i.e., without either ground control (Dial and Grodecki, 2002), bias compensation in image space (Fraser and Hanley, 2003; Fraser and Hanley, 2005; Hu et al., 2004), generic refinement (Xiong and Zhang, 2009; Xiong and Zhang, 2011) or bias-compensation in the orbital space (Teo, 2011; Teo, 2013). Additionally, as the coefficients in the simplified RFM (Equation 14) are much less than those of the regular RFM, directly refining the coefficients can be considered as an alternative usage of the simplified RFM.

As what were done earlier in this Section, 38 scenes of ALOS AVNIR2 1B2 images covering Jilin and Liaoning province of China were used to obtain a uniform simplified RFM:

$$\begin{cases} r = \dfrac{a_0 + a_1X + a_2Y + a_3Z}{1 + b_3Z} \\ c = c_0 + c_1X + c_2Y + c_3Z + c_4XY + c_{13}XZ^2 \end{cases} \tag{15}$$

The simplified RFM (Equation 15) contains five row coefficients and six column coefficients, and at least six GCPs can be used to directly refine the coefficients. As direct refinement of the coefficients is terrain-dependent, the solution is dependent on the actual terrain relief, the number and the distribution of GCPs (Tao and Hu, 2001), and commonly more than six well distributed GCPs are needed to refine the coefficients. In this test, 15 GCPs are used.

Comparative tests were carried out using a scene of ALOS AVNIR2 1B2 image (spatial resolution is 10 m), whose elevation range is from 10.35 meters to 1016.74 meters. 84 ground points were manually picked in the range of the image. The ortho-rectified ALOS PRISM images, whose geometric accuracy is better than 10 m, were used as reference images. 15 of the ground points were selected as GCPs and the rest were treated as checkpoints. In the first test, the affine transformation in image space was applied based on the provided RPCs, and in the second test, the coefficients in simplified RFM were directly corrected. The comparative result is provided in Table 4, and the distribution of the GCPs and checkpoints are shown in Figure 5.

It can be observed from Table 4 that by directly refining the coefficients in the simplified RFM, we can obtain higher accuracy than the affine-compensated RPCs. Actually, the bias compensation is finally incorporated into the originally provided RPCs (Fraser and Hanley, 2003), and it can be considered as a refinement of the original RPCs. However, directly refining the simplified coefficients can correct the coefficients to a greater extent than that can be done by the bias compensation approach, and therefore obtain higher accuracy than bias compensation when plenty GCPs are used.

Additionally, Figure 6 shows the accuracy of the affine-compensated RPCs and the directly refined coefficients using different numbers of GCPs (from 6 to 45). The GCPs were evenly selected from the 84 ground points, and the rest were

TABLE 3. THE MAXIMUM RMS AND MAXIMUM ERRORS (PIXEL) FOR 178 SCENES OF ALOS PRISM 1B2 IMAGES

Maximum of 178 scenes	Control points		Check points	
	Row	Column	Row	Column
RMS error	0.004051	0.003827	0.035605	0.007449
Maximum error	0.017200	0.012296	0.153049	0.095410

TABLE 4. COMPARISON OF THE BIAS COMPENSATION IN IMAGE SPACE AND DIRECT REFINEMENT OF THE SIMPLIFIED COEFFICIENTS; FOR EACH TEST, THE RMS AND MAXIMUM ERRORS (PIXEL) ARE REPORTED

	Bias compensation in image space				Refine the simplified coefficients			
	Control points		Check points		Control points		Check points	
	Row	Column	Row	Column	Row	Column	Row	Column
RMS error	0.43	0.56	0.45	0.57	0.32	0.48	0.44	0.54
Maximum error	0.77	-1.35	1.02	1.77	0.74	1.04	-1.04	1.7

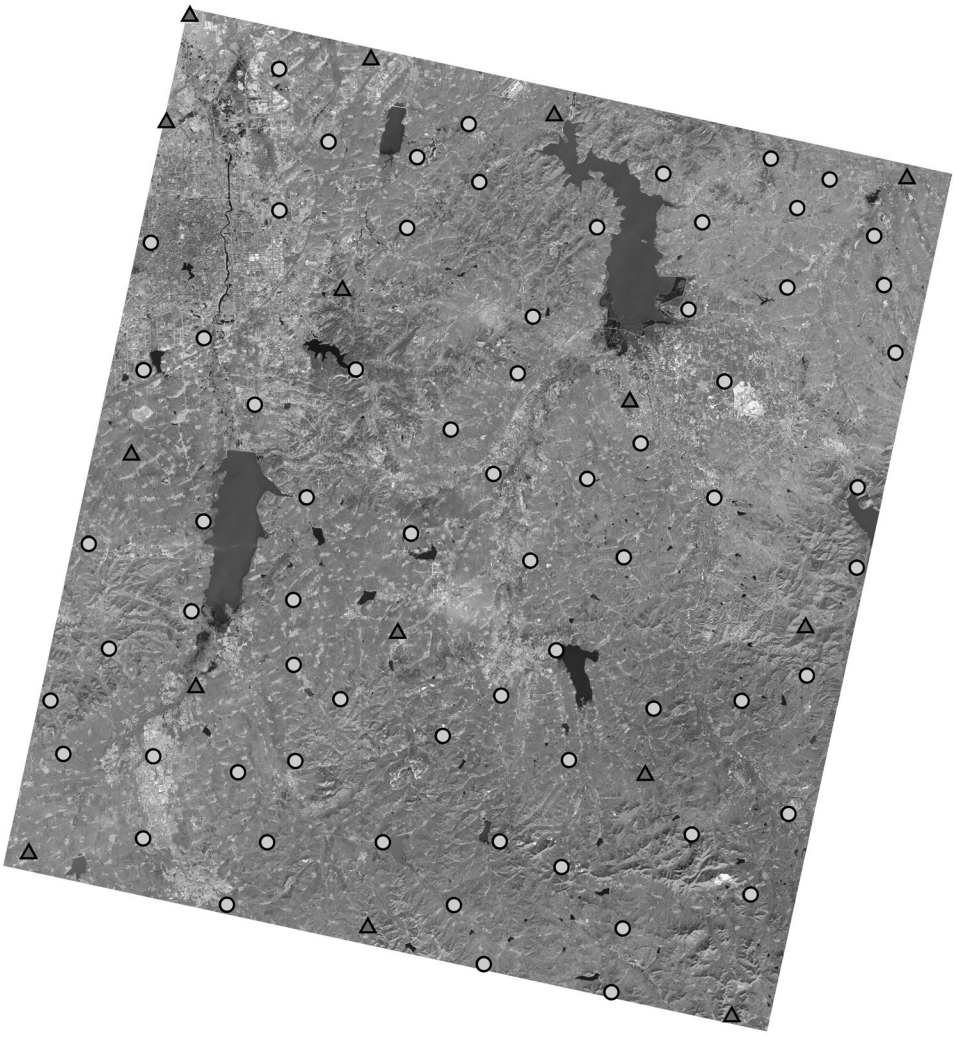

Figure 5. The distribution of the GCPs and checkpoints; GCPs are represented with triangles and checkpoints are represented with circles.

treated as checkpoints, whose RMS errors are shown in Figure 6. Obviously, the affine compensation approach is quite stable, and the accuracy does not improve much as the number of GCPs increases. On the other hand, the accuracy of direct refinement of the simplified coefficients vacillates when the number of GCPs is less than 10. However, as the number of GCPs increases, the accuracy of direct refinement becomes stable and is even higher than that of affine compensation.

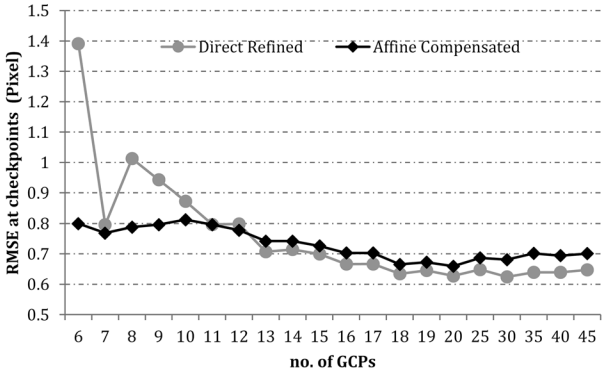

Figure 6. Accuracy comparison of two methods using different numbers of GCPs.

Bias compensation for RPCs, which is widely used terrain-independent RFM approach, requires only several GCPs or single GCP, and it is a convenient and efficient way to utilize the regular RFM as well as the simplified RFM. However, when more GCPs are available, direct refinement of the simplified coefficients can be used to further improve the correct accuracy.

Conclusions

The full RFM contains 78 RPCs, and the multicollinearity of the RPCs raises some critical problems, including the poor condition of the normal equation and rank defect of the normal matrix. In this paper, the problem of solving rational polynomial coefficients is converted into a problem of multiple linear regression, and a complete methodology of automatic selection and solution of RPCs based on nested regression is proposed. By means of analysis of the observation data (GCPs), the proposed method weeds out the unnecessary RPCs and builds a simplified model. Both problems of numerical instability and rank defect are solved by the proposed method. Compared with the conventional methods (least square regression and ridge regression), the proposed NRBOS method provides a more stable and accurate solution when the observation data are sufficient. When the observation data are not enough for the 78 unknown RPCs, the reliable solutions are not available for the conventional methods. However, the proposed NRBOS method can provide reliable solutions of RPCs using fewer than 39 GCPs. Consequently, by means of the proposed NRBOS method, terrain-dependent RFM can be applied to rectify the remotely sensed images without knowing the position and orientation information of sensor. Furthermore, when the orientation parameters of the satellite are not available or are not sufficiently accurate, the simplified RFM

shows the potential for substituting the regular terrain-independent RFM and improving the accuracy of bias-compensated RPCs.

Acknowledgments

The research has been supported by the grant from the National Natural Science Foundation of China (61271013), 135 Strategy Planning of Institute of Remote Sensing and Digital Earth, CAS and the National High Technology Research and Development Program of China (863 Program) (2012BAH27B05).

The author is grateful to the anonymous reviewers for carefully reading the manuscript and making useful suggestions.

References

Amaldi, E., and V. Kann, 1998. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, *Theoretical Computer Science*, 209(1):237–260.

Bashtian, M., M. Arashi, and S. Tabatabaey, 2011. Using improved estimation strategies to combat multicollinearity, *Journal of Statistical Computation and Simulation*, 81(12):1773–1797.

Cameron, A., and F. Windmeijer, 1997. An r-squared measure of goodness of fit for some common nonlinear regression models, *Journal of Econometrics*, 77(2):329–342.

Chong, I., and C. Jun, 2005. Performance of some variable selection methods when multicollinearity is present, *Chemometrics and Intelligent Laboratory Systems*, 78(1):103–112.

Dial, G., and J. Grodecki, 2002. Ikonos accuracy without ground control, *Proceedings of ISPRS Commission I Mid-term Symposium*, Denver, Colorado, pp. 10–15.

Draper, N., H. Smith, and E. Pownell, 1966. *Applied Regression Analysis*. Vol. 3, Wiley, New York.

Fraser, C., G. Dial, and J. Grodecki, 2006. Sensor orientation via RPCs, *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(3):182–194.

Fraser, C.S., and H.B. Hanley, 2003. Bias compensation in rational functions for Ikonos satellite Imagery, *Photogrammetric Engineering & Remote Sensing*, 69(1):53–57.

Fraser, C.S., and H.B. Hanley, 2005. Bias-compensated RPCs for sensor orientation of high-resolution satellite imagery, *Photogrammetric Engineering & Remote Sensing*, 71(8):909–915.

Geladi, P., and B. Kowalski, 1986. Partial least-squares regression: A tutorial, *Analytica Chimica Acta*, 185:1–17.

Guyon, I., and A. Elisseeff, 2003. An introduction to variable and feature selection, *The Journal of Machine Learning Research*, 3:1157–1182.

Hoerl, A., and R. Kennard, 1970. Ridge regression: Biased estimation for nonorthogonal Problems, *Technometrics*, 12(1):55–67.

Hu, Y., V. Tao, and A. Croitoru, 2004. Understanding the rational function model: Methods and Applications, *International Archives of Photogrammetry and Remote Sensing*, 20:6.

Jurczyk, T., 2012. Outlier detection under multicollinearity, *Journal of Statistical Computation and Simulation*, 82(2):261–278.

Kibria, G., and A. Saleh, 2012. Improving the estimators of the parameters of a probit regression model: A ridge regression approach, *Journal of Statistical Planning and Inference*, 142:1421–1435.

Lazraq, A., R. Cleroux, and J. Gauchi, J., 2003. Selecting both latent and explanatory variables in the PLS1 regression model, *Chemometrics and Intelligent Laboratory Systems*, 66(2):117–126.

Lin, F., 2008. Solving multicollinearity in the process of fitting regression model using the nested estimate procedure, *Quality & Quantity*, 42(3):417–426.

Lin, X., and X. Yuan, 2008. Improvement of the stability solving rational polynomial Coefficients, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, 37(Part B1):711–714.

Neter, J., M. Kutner, C. Nachtsheim, and W. Wasseman, 1996. *Applied Linear Statistical Models*, Vol. 4., Irwin, Chicago.

OGC, 1999. The open GIS abstract specification, Vol. 7: The Earth imagery case, URL: *http://www.opengeospatial.org/standards/as* (last date accessed: 05 December 2013).

Pang, H., 2011. *Econometrics*, Chengdu: Southwestern University of Finance and Economics Press.

Park, C., and Y. Yoon, 2011. Bridge regression: Adaptivity and group selection, *Journal of Statistical Planning and Inference*, 141(11):3506–3519.

Takaku, J., 2011. RPC generations on ALOS prism and AVNIR-2, *Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, IEEE, pp. 539–542.

Tao, C., and Y. Hu, 2000. Investigation of the rational function model, *Proceedings of 2000 ASPRS Annual Convention*, 22–26 May, Washington, D.C. (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), unpaginated CD-ROM.

Tao, C., and Y. Hu, 2001. A comprehensive study of the rational function model for photogrammetric processing, *Photogrammetric Engineering & Remote Sensing*, 67(12):1347–1357.

Teo, T.-A., 2011. Bias compensation in a rigorous sensor model and rational function model for high-resolution satellite images, *Photogrammetric Engineering & Remote Sensing*, 77(12):1211–1220.

Teo, T.-A., 2013. Line-based rational function model for high-resolution satellite imagery, *International Journal of Remote Sensing*, 34(4):1355–1372.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society*, Series B (Methodological), pp. 267–288.

Tobias, R., 1995. An introduction to partial least squares regression, *Proceedings of the Annual SAS Users Group International Conference*, Orlando, Florida, pp. 2–5.

Wang, H., 1999. *Partial Least Squares Regression Method and its Application*, Beijing: National Defense Industry Press.

Xiong, Z., and Y. Zhang, 2009. A generic method for RPC refinement using ground control Information, *Photogrammetric Engineering & Remote Sensing*, 75(9):1083–1092.

Xiong, Z., and Y. Zhang, 2011. Bundle adjustment with rational polynomial camera models based on generic method, *IEEE Transactions on Geoscience and Remote Sensing*, 49(1):190–202.

Yuan, X., and J. Cao, 2011. An optimized method for selecting rational polynomial coefficients based on multicollinearity analysis, *Geomatics and Information Science of Wuhan University*, 36(6):664–669.

Yuan, X., and X. Lin, 2008. A method for solving rational polynomial coefficients based on ridge estimation, *Geomatics and Information Science of Wuhan University*, 33(11):1130–1133.

Zhang, Y., Y. Lu, L. Wang, and X. Huang, 2012. A new approach on optimization of the rational function model of high-resolution satellite imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 50(7):2758-2764.

(Received 08 May 2013; accepted 21 August 2013; final version 08 October 2013)